

604184

604184

1

THE VALIDITY OF PEER RATINGS FROM OCS IN PREDICTING OFFICER PERFORMANCE:

A FOLLOW-UP OF THE 1955 NEWPORT STUDY

E. P. Hollander

State University of New York at Buffalo

COPY <u>1</u> OF <u>1</u>	
HARD COPY	\$ . 1.00
MICROFICHE	\$ . 0.50

10p

Technical Report

May 1964

Prepared under ONR Contract 816(12)

✓ h g

Department of Psychology

Washington University

St. Louis, Missouri

Reproduction in whole or in part is permitted  
for any purpose of the United States Government.

DDC  
AUG 21 1964  
DDC-IRA C

ch

## Summary

This report covers the follow-up phase of a study of peer nominations begun in 1955 at the Naval Officer Candidate School in Newport, Rhode Island. <sup>as reported</sup> ~~1000~~ 700 trainees, in the 23 sections comprising an entire OCS class, completed several peer nomination forms at various stages of training. One form, in particular, requiring nomination for "success as a future Naval Officer," ~~1000~~ was administered to all sections.

From among the trainees in the original study, 639 were identified who had gone on to duty as officers for a period averaging three years or more. Fitness report ratings given to these officers by their seniors were obtained and averaged to yield a score representing the performance criterion. This score had a corrected split-half reliability of .90.

In the prediction of this officer performance criterion, the FO peer nomination score from the third week of training yielded an average validity of .40. This value was not exceeded for the peer nomination scores secured from later administrations of the same form, thus demonstrating the success of this early evaluation in predicting later performance as an officer.

The only predictor from OCS achieving a level of validity approximating that for the third week peer nomination was the final academic grade which had a correlation of .41 with this performance criterion. A partial  $r$  for the correlation of third week FO with the criterion, holding the academic grade constant, gave a value of .28. Since a score for frequency of choice as a friend was found to correlate .22 with the criterion, another partial  $r$  was calculated for third week FO with performance as an officer this time holding friendship constant. The resulting value was .33. In both instances, though the validity coefficients were diminished, the unique contribution to prediction provided by the third week FO peer nomination was maintained at a respectable and useful level.

## Major Conclusions

The validity of early peer nominations, already established for performance in OCS, was ~~also~~ found to be substantial for the prediction of later performance as an officer. Validity is maintained even after statistical control for academic performance and popularity is applied by the partialing process. It is therefore recommended that early peer nominations, assessing overall performance as an officer, be routinely utilized early in training as a supplement to academic and other evaluations. ( )

\*The author is indebted to the officers and trainees of the OCS in providing a cooperative setting for the original study, and to Sidney Friedman, Victor Fields, and Joseph Cowan of DuPers, as well as Luigi Petrullo and Abraham Levine of ONR, for considerably facilitating the follow-up phase of this work. Leonore Ganschow and Karol Anderson were of inestimable aid in helping with the analyses and the preparation of this report.

## I. Introduction

One of the more persistent problems in the assessment of trainees is the early identification of those who are unlikely to meet performance requirements. Whether these exist in terms of the training itself, or the later demands of the job, this issue is of considerable concern. The earlier such information is available, the greater its utility in realizing an appreciable savings in time, effort, effectiveness, and other costs. Moreover, special value is attached to measures which can provide additional information on which to base evaluations.

Among those measures which supplement the more usual kinds of grades in training programs is the peer rating. This involves each group member's assessment of every other group member on a recognizable quality such as leadership, popularity, or performance. Thus, a peer rating represents an evaluation made by those who occupy comparable status and who have an immediacy of contact with one another over time. From these ratings, a composite score may be obtained which can be used to predict the criterion or serve itself as a criterion for validation. There are two essential variations of this process: ranking, in which each member ranks or assigns a score to his peers; and nominating, in which a specified number of group members are named as "high" or "low" on the quality being assessed.<sup>1</sup>

Because it is more manageable than either the ranking or scaling procedures noted, the most common rating form is the peer nomination; furthermore, peer nominations have been found to yield substantial validity and reliability (cf. Hollander, 1954, 1964). Scores from these "pooled group judgments" contribute unique variance which might not otherwise be available in the evaluation of trainees and in the prediction of their later performance. In particular, this report presents data on the validity of peer nominations gathered in OCS on "success as a future Naval Officer."

## II. Background

In 1955, under ONR Contract 760(06), the author undertook a study to answer several questions regarding the optimum utilization of peer nomination procedures. With the cooperation of the Naval Officer Candidate School at Newport, Rhode Island, an entire OCS class, composed of 23 trainee sections, was made available for this purpose.

The aim of this basic study was to assess the effects on the reliability and validity of nomination scores arising from four factors: the length of time the members of the section had been together; the use of a "research" set vs. a "real" set on the peer nomination forms; the nature of the quality to be rated; and the effects of friendship choice. The specific findings of the study are contained in several sources in the bibliography, but a general summary is given in Hollander (1956d). Some of these findings can be detailed briefly here.

Four forms, specifying different qualities to be rated, were utilized. They dealt with leadership, motivation for naval service, probability of success in OCS training, and success as a future officer. All forms were administered at least three times during the training period of sixteen weeks' duration, usually at the time of initial orientation, and again at the third and the sixth week of training.

---

<sup>1</sup>A fuller discussion of the character and application of peer nominations will be found in E. P. Hollander, Leaders, Groups, and Influence, New York: Oxford University Press, 1964, notably Chapter 8.

The reliability of the forms was calculated at each time of administration by the split-half method. It was found that even the earliest nomination scores, derived from forms administered after four or five days of contact (the "0" Week) yielded reliability coefficients of about .90. The increase in reliability was not appreciable for the later scores. Furthermore, a high repeat reliability was found for the same form administered at different times.

Validity was initially determined from criteria accessible in the training program, i.e. pass-fail, final academic average, and final military aptitude grade assigned by superiors. For these criteria, the peer nomination scores gave significant, and differentially discriminating, validity coefficients. As an example, nominations for probability of success in the school, secured before the onset of formal classes, were significantly correlated with the ultimate pass-fail and academic criteria. In general, nomination scores obtained very early in training, and certainly by the third week, provided completely comparable validity to that of later nominations. Though different forms gave significant differences in validity against various criteria, there was no general disparity between the validity of forms administered under the "research" as against the "real" set.

As previously reported by others (e.g., Wherry & Fryer, 1949), the popularity dimension represented in friendship choice was not found to have a major intrusive effect on the validity of these nominations. A "friendship score," based upon the number of friendship choices received by a section member, was found to be significantly correlated with peer nomination scores, but this score was not systematically related to the criterion of academic performance. When validity coefficients for peer nominations on "success in OCS" were corrected by partialing friendship, it was found that the validity level was retained. Furthermore, different forms yielded different relationships with friendship, indicating the operation of a selective factor and not a general bias.

Although these findings illuminated paths of application, it was essential that additional data be obtained for the construction of a post-training criterion against which peer nomination forms might be further validated. This report presents a follow-up study in that vein.

### III. Procedure

All of the more than 700 trainees available at the outset of this study were given a primary form calling for nominations on "success as a future Naval Officer" (FO). This was seen to be of particular worth in its likely prediction of more distant, officer performance criteria. In addition to this primary form each section received one of three so-called secondary forms, i.e., "leadership qualities" (LQ), "interest in and enthusiasm for the Naval Service" (IE), and "probability of success in OCS" (OC). As has been previously indicated, these forms had differential validity in predicting the in-training criterion, particularly academic performance, with the OC form being the highest in this specific regard.

Cutting across this pattern, approximately half the sections received a "research" set (RO) with the explicit point, appearing on their peer nomination forms, that the results of the ratings were to be used for research purposes only. The other sections were given equally explicit instructions that the



results might be used administratively (AU). This split in treatment, designed to provide data on differential reliability and validity, gave only minor differences mainly in terms of form-set interactions.

In Appendix A, Form FO-AU is reproduced as an illustration of the format followed with all forms. Since this was varied only slightly to accommodate alternative instructions, the reader may view this as an example of the general form applied. As will be noted, the form requires five "high" and five "low" nominations in order of preference. Each of the subjects was provided with a complete alphabetical roster of the section mates every time he was required to complete a form. The author was the sole administrator of the forms for all sections at all times.

A direct weighting procedure was applied to derive peer nomination scores. The highest nominee was awarded a +5, the next highest, a +4, and so on through the 5 "highs"; similarly, the lowest nominee was assigned a -5, the next lowest, a -4, and so on. An algebraic sum was then obtained for each subject and divided by the N of the group minus 1, since no subject could nominate himself. This resulted in an average score ranging on a continuum from +5 to -5. To remove the minus sign, a constant of 5 was added to this score and the resultant value was then multiplied by 10 in order to permit the use of a two-digit score without the intervening decimal point. The distribution arising from this procedure has normal characteristics with a mean of about 50 and a standard deviation approximating 10 for the total population of the study. Though this score may be seen to have certain features of the standard score, it neither obscures section differences, as does the standard score, nor does it presume homogeneous characteristics from section to section.

The criterion of officer performance applied in the follow-up phase of this study was derived from question 16a of the Standard Report on the Fitness of Naval Officers (NAVPERS-310, revised 3-54). This question is given below with the weights we used in calculating an average score:

"In comparison with other officers of his grade and approximate length of service, how would you designate this officer?"

- 5 One of the few highly outstanding officers I know
- 4 A very fine officer of great value to the service
- 3 A dependable and typically effective officer
- 2 An acceptable officer
- 1 Unsatisfactory (adverse)

The 639 officers in the follow-up study had been in active service for an average of no less than 3 years during which time they were usually evaluated twice yearly. The distribution of average scores received on this item of the "fitness report" was essentially unimodal with a mean of about 3.6 and a standard deviation of about .5. A split-half reliability analysis of this rating yielded an uncorrected score of .81 which reached .90 when corrected by the Spearman-Brown formula.

This question also has been found to have a high intercorrelation with other scales on this fitness report (see King & Wollack, 1960, p. C4) especially with

the assessment of various qualities, notably leadership, in question 19. The distribution of individual ratings received by these junior officers on the whole was slightly lower than those found for the representative sample of ensigns and lieutenants junior grade in the study just cited. This may be accountable in terms of the inclusion in their sample of Regular Navy Officers whereas our sample was made up entirely of Reserve Officers.

#### IV. Results

Table 1 summarizes the validity coefficients obtained for the four peer nominations administered at three time periods. As will be seen there, "future officer" peer nomination is the best predictor among them of the criterion of officer performance obtained from fitness reports over three years or more. However, the other peer nomination forms quite generally give substantial and significant prediction, beginning with the third week, for both the in-training and the post-training criteria. In view of the high average correlation (.90) between the FO and LQ forms, it is no surprise that LQ should so closely approximate the validity of FO.

Taking account of the correlation of .40 between the third week FO peer nomination and the officer performance criterion, it is evident that even at this early level of exposure to one another, prediction of the more distant criterion is substantial. Table 2 provides a matrix of intercorrelations for the FO peer nominations, other OCS predictors, and the officer performance criterion. It is noteworthy that the third week FO validity is as high as that obtained in the administration at the twelfth week, and furthermore that the only other variable reaching this level of validity is the final OCS academic grade, which correlates .41 with the officer performance criterion. In view of the fact that the OCS academic grade and the third week FO score correlate .42, a partial  $r$  was computed holding academic performance constant; the resulting validity of the third week FO score in predicting the officer performance criterion, with academics partialled, was .28.

In another line of analysis, the third week FO score, the final academic grade, and the OCS Mathematics score were combined into a multiple correlation with the officer performance criterion. This procedure gave an  $R$  of .51, with the beta weights for FO and final academic grade being high and of essentially the same magnitude, .29.

Still another way of viewing the predictive effectiveness of the third week FO peer nomination is to compare trainees who are in the upper segment of the distribution on FO with those in the middle and in the lower part of the distribution. In pursuing this, 6 sections of the entire sample, a total  $N$  of 174, were divided into those who had reached a score of 56 or over (.6 sigma above the mean) on the third week FO, those who had scored between 55 and 45 (.5 sigma above to .5 sigma below) on that measure, and a bottom segment scoring 44 or under (.6 sigma below). Those in the upper group (28%) had an average fleet performance score of 3.90, those in the middle (46%) a score of 3.63, and those in the lower group (26%) a score of 3.36; all of these differences are significant at or beyond the .05 level. This reflects the early discriminatory power the FO form has in predicting thresholds of later officer performance.

Table 1

Average Validity Coefficients\* against Three Performance Criteria  
for Four Peer Nomination Scores from Three Stages of Training

<u>Peer Nominations</u>		<u>Criteria</u>								
Week Completed		Final OCS Military Grade			Final OCS Academic Grade			Officer Fitness Report		
		<u>0</u>	<u>3</u>	<u>6</u>	<u>0</u>	<u>3</u>	<u>6</u>	<u>0</u>	<u>3</u>	<u>6</u>
"Future Officer" (FO)		37	45	46	14	42	45	24	40	39
N = 639										
"Interest and Enthusiasm" (IE)		40	45	40	17	32	23	19	33	33
N = 228										
"Success in OCS" (OC)		33	38	35	31	72	77	26	36	35
N = 182										
"Leadership Qualities" (LQ)		41	47	51	19	33	39	26	36	37
N = 229										

\*All decimal points removed.

The total number of friendship choices received by a trainee had previously been found not to correlate with the OCS academic criterion, but the third week friendship score in this follow-up was found to correlate with the performance criterion at a level of .22. The repeat reliability of the third week friendship score was .82. When the validity coefficient of .40 for the FO third week is corrected by a partial  $r$  taking account of friendship's correlation of .58 with FO, this gives a corrected validity coefficient of .33.

Unlike the study by Wollack and Guttman (1961) no distinction was made in this study between officer assignments to shore and fleet billets. Nonetheless, the correlation we obtained of .40 accords well with the .33 validity coefficient obtained by them for peer ratings at the eleventh week of OCS correlated with the same scale from the fitness report for just fleet billets. We did, however, endeavor to determine the predictability of the third week FO score for line officers only, with a sub-sample of 135 line officers drawn at random. This yielded an average correlation of .54 which was substantially greater than the  $r$  of .40 obtained for the entire sample with the criterion. Moreover, our results fit those reported by Weitz (1958), who found a relationship of .40 between peer ratings and later supervisory ratings for life insurance agents in a quite different sphere.

Table 2

Matrix of Intercorrelations\* of OCS Predictors  
and Officer Performance Criterion

N = 639

	2	3	4	5	6	7	8	9	10	11	12
1 FO Peer Nomination "0" Week	75	65	56	03	07	-06	11	08	37	14	24
2 FO Peer Nomination "3" Week		92	81	22	21	18	25	17	45	42	40
3 FO Peer Nomination "6" Week			89	21	21	22	29	17	46	45	39
4 FO Peer Nomination "12" Week				16	22	21	28	16	45	43	40
5 OCB Verbal Reasoning					29	34	28	23	08	41	15
6 OCB Mechanical Comprehension						50	42	56	14	38	13
7 OCB Mathematics							43	37	13	57	22
8 OCB Relative Movement								38	16	39	15
9 OCB Spatial Relations									12	33	11
10 Final OCS Military Grade										46	26
11 Final OCS Academic Grade											41
12 Officer Fitness Report Criterion											

\*All decimal points removed.

As a final point, we found that in almost every case, following graduation, the new officers had gone on to duty under instruction. Modally, they had received one rating each while in this status. An analysis to see the effect of this duty under instruction rating upon the overall officer performance rating indicated that it had no significant effect either upon the overall reliability of the officer performance score or upon the validity of the FO peer nomination in predicting this criterion.

Whatever mode of analysis, it seems apparent therefore that early peer nominations can contribute unique variance in the prediction of criteria generated long after training. This finding strongly encourages the continued use of such evaluations early in training as a supplement to other measures.



## Bibliography

- Hollander, E. P. Buddy ratings: Military research and industrial implications. Personnel Psychol., 1954, 7, 335-393.
- Hollander, E. P. Conditions affecting the military utilization of peer ratings: The Newport study. I. Reliability. Navy Tech. Rep. 1-56. Pittsburgh: Psychological Laboratories, Carnegie Institute of Technology, January, 1956. (a)
- Hollander, E. P. Conditions affecting the military utilization of peer ratings: The Newport study. II. Validity against in-training criteria. Navy Tech. Rep. 2-56. Pittsburgh: Psychological Laboratories, Carnegie Institute of Technology, February, 1956. (b)
- Hollander, E. P. Conditions affecting the military utilization of peer ratings: The Newport study. III. Friendship choice. Navy Tech. Rep. 3-56. Pittsburgh: Psychological Laboratories, Carnegie Institute of Technology, April, 1956. (c)
- Hollander, E. P. Conditions affecting the military utilization of peer ratings: The Newport study. Final Report, ONR Contract 760(06). Pittsburgh: Psychological Laboratories, Carnegie Institute of Technology, May, 1956. (d)
- Hollander, E. P. The friendship factor in peer nominations. Personnel Psychol., 1956, 9, 435-447. (e)
- Hollander, E. P. Interpersonal exposure time as a determinant of the predictive utility of peer ratings. Psychol. Reps., 1956, 2, 445-448. (f)
- Hollander, E. P. The reliability of peer nominations under various conditions of administration. J. appl. Psychol., 1957, 41, 85-90.
- Hollander, E. P. Leaders, groups, and influence. New York: Oxford Univer. Press, 1964.
- King, D. T., & Wollack, L. Study of officer fitness reports submitted in 1957 and early 1958. Pers-152 Memo 60-10, Personnel Research Division, BuPers, September, 1960.
- Weitz, J. Selecting supervisors with peer ratings. Personnel Psychol., 1958, 11, 25-35.
- Wherry, R. J., & Fryer, D. H. Buddy ratings: Popularity contest or leadership criterion? Personnel Psychol., 1949, 2, 147-159.
- Wollack, L., & Guttman, I. Prediction of OCS academic grades and post-OCS performance of junior officers with a battery of speeded tests. Bureau of Naval Personnel, Tech. Bull. 61-2, January, 1961.

## APPENDIX A--FORM FO-AU

U. S. NAVAL SCHOOL, OFFICER CANDIDATE  
NEWPORT, RHODE ISLAND

You have been together with the men in this room since you entered OCS. From this contact, you will have formed certain impressions of them regarding their future success as Naval Officers.

Considering these impressions, and carefully weighing the qualities required in a successful Naval Officer, you are to select the five members of your section whom you consider to have the highest promise as Naval Officers, and the five members of your section whom you consider to have the lowest promise as Naval Officers.

THE RESULTS OF THESE RATINGS MAY BE USED FOR ADMINISTRATIVE PURPOSES.

Specifically, you are to perform these exact operations:

1. Consult the section roster which has been provided you and draw a line through your own name.
2. Study the remaining names on the roster and select the individual with the highest promise of success as a Naval Officer. Enter this name in the space labeled H-1 below and then draw a line through that name on the roster.
3. Study the roster again and then select the individual with the lowest promise. Enter this name in the space labeled L-1 below and then draw a line through that name on the roster.
4. Continue the study of this roster, alternately selecting individuals with the highest promise and the lowest promise, until you have entered ten names. Draw a line through each name on the roster as you write it in the proper place here.

HIGHEST PROMISE \_\_\_\_\_ H-1

\_\_\_\_\_ H-2

\_\_\_\_\_ H-3

\_\_\_\_\_ H-4

\_\_\_\_\_ H-5

\_\_\_\_\_ L-5

\_\_\_\_\_ L-4

\_\_\_\_\_ L-3

\_\_\_\_\_ L-2

LOWEST PROMISE \_\_\_\_\_ L-1